
*
* **FINAL REPORT:** *
*
* **TIMESHARING PERFORMANCE** *
*
* **AS AN** *
*
* **INDICATOR OF PILOT MENTAL WORKLOAD** *
*

Prepared by

Patricia A. Casper
Department of Psychological Sciences
Purdue University
West Lafayette, IN 47907

for

NASA-Ames Research Center
Moffett Field, CA 94035
through
Division of Sponsored Programs
Purdue Research Foundation
West Lafayette, IN 47907

Project Number: NCC 2-349

PRINCIPAL INVESTIGATOR(S)

BARRY H. KANTOWITZ January 1, 1985 - August 15, 1987

ROBERT D. SORKIN August 15, 1987 - May 31, 1988

(NASA-CR-185328) TIMESHARING PERFORMANCE AS
AN INDICATOR OF PILOT MENTAL WORKLOAD Final
Report (Purdue Univ.) 31 p CSCL 05H

N89-25573

Unclas
G3/54 0217269

I. Summary of completed research (1/1/85 - 12/31/86)

The sponsored research was carried out in two simultaneous phases, each intended to identify and manipulate factors related to operator mental workload. The first phase concerned evaluation of attentional deficits (workload) in a timesharing task developed at the human information processing laboratory at Purdue. Work in the second phase involved incorporating the results from these and other experiments into an expert system designed to provide workload metric selection advice to non-experts in the field interested in operator workload. For the most part, the results of the experiments conducted are summarized in general, with the details available in the papers found in the Appendices.

Two years of research at Purdue and at NASA-Ames were successful in identifying some of the salient factors associated with operator mental workload in complex task situations. In the laboratory at Purdue, a series of experiments using a bimodal (auditory and visual) divided attention task has revealed that operators are not limited in their ability to attend to simultaneous events - the limitations arise when they are required to make responses to them. Cross-sectioned slices of the task's data at different points in time showed that performance in one modality is not affected by perceptual events in the other modality, such as changes in color of a display, for example. Performance on the task in one modality suffered the most when something occurred in the other modality that required a response, such as the flash of a light that had special significance to the operator (Casper, 1986; Kantowitz & Casper, in press).

The ratio of sub-tasks also appears to be an important factor affecting workload. According to previous research supported by Gestalt principles of perception, (Klapp, Hill, and Tyler, 1983), tasks in which the ratio of visual to auditory events is a harmonic one (e.g. 1:1, 2:1) should be easier than tasks in which there is a "non-harmonic" ratio between the stimuli (e.g. 3:2). Our research, based on attention theory, found the opposite to be true -- tasks having a 3:2 ratio of visual to auditory stimuli are in fact easier to perform than tasks in which the auditory and visual events occur simultaneously (Casper & Kantowitz, 1985). In some respect this advantage was due to the fact that there was more asynchrony of processing demands between the two tasks in the 3:2 case. This hypothesis is supported by the results of an experiment in which there was some advantage when two tasks of the same ratio were presented with one task lagging slightly behind the other.

The second phase of the research focused on the transfer of workload knowledge obtained from empirical studies to the practitioner responsible for making workload-related decisions.

With the ever-increasing concern for issues of safety, efficiency, and enhanced performance in aviation, more and more people are becoming interested in pilot workload. However, not all those interested in workload have adequate knowledge about it to know how to begin evaluation. Recent advances in the field of artificial intelligence have made several expert system development tools available to those who wish to build expert or decision support systems but who haven't the programming knowledge (or money) themselves to create it from scratch. In essence, the tools are "expert systems for building expert systems".

During the summer of 1986, one of these tools was used to create a microprocessor-based prototype of a system that makes recommendations concerning the choice of workload metrics, depending on the user's research goals, available equipment, and task environment (Casper, Shively, & Hart, 1986). After asking the user a series of questions related to these factors, the thirteen workload metrics are ordered with respect to their appropriateness for the user's situation. Metrics that are completely inappropriate for the user are not suggested at all, while more appropriate measures are offered with a number from 1 to 10 indicating the degree to which they would be helpful to the user. Following the suggestions, the program allows the user to access workload information files where he may learn more about any of the measures included in the database. The files represent the most current available information concerning a measure's empirical success, its practical limitations, and its advantages and disadvantages. These files may be viewed on the computer screen or routed to a printer to obtain a hard copy.

II. Summary of completed research during extension (1/1/87 - 5/31/88)

A. Laboratory research at Purdue

The research conducted at Purdue during the final year of the project addressed two questions:

- 1) Is heart rate variability a valid indicator of overall operator mental workload in our laboratory task?
- 2) If it is, can it be used to provide further insight on the location of bottlenecks in the human information processing system?

In general, three broad classes of measures have been used to assess operator workload: subjective, performance, and physiological. Not surprisingly, each class of measures (and even measures within classes) have their unique advantages and disadvantages. Unlike the first two classes of measures, the physiological indices of workload have a distinct advantage in that they are unobtrusive to the operator performing a task.

Although not as crucial a factor in the laboratory it is obvious why unobtrusiveness would be advantageous in an operational situation such as a low-altitude helicopter mission, for example.

Much research has been done supporting the use of various computations of heart rate variability as an indicator of operator workload, ranging from the standard deviation of the variability of the interval between successive heart beats to a more complicated analysis of the power in different spectral bands of the heart beat signal. A review of the current literature has raised several issues, namely, the problem of contamination of overall heart rate variability with influences from factors other than mental load, how to reconcile conflicting results obtained from different computations of variability, and the problem of how to interpret heart rate variability changes within the context of a given experimental design (Kalsbeek, 1973).

An experiment using the aforementioned divided attention task was conducted at Purdue in order to evaluate the sensitivity of cardiac measures of workload to manipulations of the difficulty of the task. Subjects simultaneously attended to two streams of discrete stimuli, and responded manually to changes in one modality and vocally to changes in the other modality. The events in the auditory modality were high or low-frequency tones and the visual events were flashes of a red or green light. Subjects were instructed to respond as quickly as possible via either a keypress or by saying the word "diff" into a microphone each time they observed a signal in a modality that was different from the previous signal in that modality. Half of the subjects used a vocal response to the auditory channel and a manual response to the visual channel, while for the other half of the subjects the response requirements were reversed. It should be noted that the response mappings for the former group should lead to better performance, since input and output modalities are more compatible for the auditory task than those used by the latter group (Wickens, 1980). Tasks employing multiple modalities are useful in that they parallel tasks in operational environments more than the traditional laboratory tasks, both in their difficulty and in their multimodal nature.

Task difficulty was manipulated by varying the number of tasks simultaneously performed (one = single stimulation, two = double stimulation) and the degree of synchrony between two tasks. In the synchronous case, the auditory and visual stimuli occurred simultaneously, and in the asynchronous case, presentation of the auditory or the visual sequence was delayed by 300 msec after that in the other modality. Presumably, tasks that occur asynchronously in each modality are easier to perform since attention may be switched between the two and responses need not necessarily be executed simultaneously. Due to equipment malfunctions heart rate (HR) data was available for only 6 of the 24 subjects run in the experiment. Mean HR scores showed that HR decreased throughout the experiment, presumably

indicating decreased arousal throughout the experiment. Of HR, HR variance, mean successive difference in IBI's (MSD), and variance of successive differences in IBI's, only mean HR reflected differences between the pre-task baseline period (82 beats per minute [BPM]) and the task period (76 BPM). HR did not distinguish between the single and double stimulation versions of the task.

HR variance was significantly greater during the last half of the experiment than in the first half, but decreased within a half, perhaps reflecting the fact that subjects were growing increasingly fatigued and exerting greater effort during the portions of the experiment between rest periods. The experiment lasted for an hour and a half. Spectral analysis of the IBI data did not reveal differential sensitivity of four different frequency bands to the manipulations of task difficulty.

The results of the performance measures are summarized in the paper appearing in the Appendix (Casper & Kantowitz, 1987).

There are a number of reasons why the cardiac measures were not sensitive to difficulty manipulations that have previously been successful as measured by performance. The first reason has to do with the motivational state of the subjects. The subjects used for the experiment were students in an introductory psychology class and received 1 1/2 hours of course credit simply for showing up for the experiment. No performance criteria were imposed on the subjects. It is possible that the task did not cause differing degrees of effort in the subjects. The task was purposely made difficult in order to elevate the level of missed responses and false alarms needed for the signal detection measure of performance. In addition, the subjects were not given explicit performance feedback. After a block of trials the experimenter simply told them whether or not they had achieved the 50% criterion necessary for remaining in the experiment. Other studies have demonstrated that feedback is an essential component in the operator-task loop. Second, it is possible that the requirement of a vocal response to one of the tasks forced the subjects to use an artificial breathing pattern in order to keep up with the regularly-paced task. Previous studies have shown that changes in breathing patterns are reflected in HR data, possibly obscuring the effects of other experimental factors. Further, it is entirely possible that the null results reflect insufficient power in the design, since 3/4 of the HR data was not available due to the equipment malfunction. However, one would at least expect a trend in the right direction.

If cardiac measures are to be successfully used in future timesharing experiments like those described above, some modifications to the task should be made. Vocal responses are probably to be avoided in tasks where breathing is likely to be entrained to a regular rhythm. Further, motivational incentives are likely to induce the subjects to invest a greater degree of

effort and involvement in the task. The subjects could be offered a cash prize for attaining a given score on the task, and frequent feedback could be used to inform the subject of his progress toward meeting those goals.

In addition to the methodological problems mentioned above, the study of cardiac correlates of cognition is sorely in need of some kind of theoretical framework from which empirical data can be predicted and interpreted. It is likely that progress toward such a goal will require that the concept of mental workload be viewed as a multidimensional construct, with different classes of measures tapping different dimensions of workload.

B. Expert system development at NASA-Ames and Purdue

The prototype of the system created during the summer of 1986 was extensively revised and tested using several different testing methods. Some of the workload measures from the original prototype were dropped, and some new ones were added, mostly a number of secondary tasks and rating scales. The questions asked of the user were also extensively revised; useless questions (those that did not warrant a distinction among workload measures) were dropped, and the wording on ambiguous questions was clarified. In addition, if secondary task measures were among the measures suggested to the user, potential input and output modalities for those tasks were determined.

One of several major problems currently facing the field of artificial intelligence is how to qualitatively and quantitatively evaluate the validity of the advice or information provided by intelligent decision aiding systems. No standard methods of testing exist, either across or within fields of application. Since WC FIELDE is potentially a very useful and attractive tool for many researchers, several attempts at mathematical validation have been made. Although the absolute utility of the advice WC FIELDE provides will ultimately be revealed by the user, several methods of assessing its overall and relative sensitivity have been devised.

The first evaluation was used to determine the baseline sensitivity of the system to the user's input. Random numbers were used as input to the system, on the assumption that if the probabilities associated with the rules had more influence on the results than the user's input then the output of the system would always be the same. If the system is truly sensitive to the answers the user supplies to the questions then the results should be different each time the system is run. The mean correlation among 20 random runs for version 1.0 of WC FIELDE was $r = .42$, and for version 2.0 it was $r = .16$. Thus, revisions to the system were effective in increasing the sensitivity of the system to the user's input.

A second sensitivity test was performed in which hypothetical users answered the questions posed by WC FIELDE as if they were the experimenter seeking workload advice for operators performing both easy and difficult tasks in two different environments. Ten subjects were asked to answer the questions for: 1) driving a car with a manual transmission in city traffic, 2) driving a car with a manual transmission on the freeway, 3) taking a calculus final, and 4) taking an introductory psychology final. No other information was given to the subjects. One group of ten subjects used version 1.0 of WC FIELDE, while another ten subjects used version 2.0. The correlation among the measures recommended by the program for all subjects using each of the four conditions was calculated.

The correlations among measures was higher within than between environments, suggesting that WC FIELDE was able to recommend different kinds of measures depending on the task environment or situation of the experimenter. In addition, the correlation within levels of difficulty was higher than between levels of difficulty for the final exam task in version 2.0, but for neither tasks in version 1.0. That is, the kinds of measures recommended for a situation in which the operator is performing an easy task are different from those suggested for when the operator is performing a more difficult task. The fact that WC FIELDE is able to make this distinction is significant, since one of the most important factors in determining which class of measures will be most sensitive to workload variations is the level of workload the operator is expected to withstand. Future testing of the system will employ more distinct difficulty and environment manipulations in order to allow the correlations to fully work out.

As WC FIELDE is revised, it will continue to be tested using the methods outlined above. In addition, several more novel testing methods have been suggested, which should provide converging evidence of the validity of ongoing revisions. The system is currently being distributed to those who request it, so that it can begin to help those who want to assess workload, and so that user feedback can guide future system revisions.

Although decision support systems are not intended to replace the humans after which they are modeled, it is hoped that systems such as WC FIELDE will encourage increased workload evaluation in the early stages of man-machine system design, to ensure greater economy, safety, and productivity. Also, as more knowledge about workload is incorporated into the system we should come closer to a true understanding of the nature of workload and the many factors affecting how it is expressed in the human operator.

ORIGINAL PAGE IS
OF POOR QUALITY

REFERENCES

Papers marked with an asterisk were prepared under the support of this contract and are included in their entirety in the appendices of this report.

- Casper, P.A. (1986) A signal detection analysis of bimodal attention: Support for response interference. Unpublished * master's thesis, Purdue University.
- Casper, P.A., & Kantowitz, B.H. (1987) Estimating the cost of mental loading in a bimodal divided-attention task:
* Combining reaction time, heart-rate variability and signal detection theory. Paper presented at the 1987 Mental State Estimation Workshop sponsored by NASA-Langley Research Center, Williamsburg, VA.
- Casper, P.A., & Kantowitz, B.H. (1985) Seeing tones and hearing rectangles: Attending to simultaneous auditory and visual * events. In R.E. Eberts & C.G. Eberts (Eds.), Trends in Ergonomics/Human Factors II. Amsterdam: North-Holland.
- Casper, P.A., Shively, R.J., & Hart, S.G. (1987) Decision support for workload assessment: Introducing WC FIELDE.
* Proceedings of the Human Factors Society 31st Annual Meeting, 72-76.
- Casper, P.A., Shively, R.J., & Hart, S.G. (1986) Workload consultant: A microprocessor-based system for selecting * workload assessment procedures. In Proceedings of the 1986 IEEE International Conference on Systems, Man, and Cybernetics, 1054-1059.
- Kalsbeek, J.W.H. (1973b) Do you believe in sinus arrhythmia? Ergonomics, 16, 99-104.
- Kantowitz, B.H., & Casper, P.A. (in press) Mental workload in aviation. In E.L. Weiner & D.C. Nagel, (Eds.), Human Factors in Modern Aviation.
- Klapp, S.T., Hill, M.D. & Tyler, J.G. (1983) Temporal compatibility in dual motor tasks: A perceptual rather than motor interpretation. Paper presented at the 24th Annual Meeting of the Psychonomic Society, San Diego, CA.
- Wickens, C.D. (1980) The structure of attentional resources. In R.S. Nickerson (Ed.), Attention and Performance VIII, Hillsdale, N.J.: Erlbaum.

APPENDIX A:

Casper, P.A., & Kantowitz, B.H. (1985) Seeing tones and hearing rectangles: Attending to simultaneous auditory and visual events. In R.E. Eberts & C.G. Eberts (Eds.), Trends in Ergonomics/Human Factors II. Amsterdam: North-Holland.

APPENDIX C:

Casper, P.A., & Kantowitz, B.H. (1987) Estimating the cost of mental loading in a bimodal divided-attention task: Combining reaction time, heart-rate variability and signal detection theory. Paper presented at the 1987 Mental State Estimation Workshop sponsored by NASA-Langley Research Center, Williamsburg, VA.

Estimating the cost of mental loading in a bimodal divided-
attention task: Combining reaction time, heart-rate variability
and signal-detection theory

Patricia A. Casper and Barry H. Kantowitz
Purdue University
West Lafayette, IN

The topic of workload has drawn considerable interest in the field of ergonomics for a number of years. For as long as man has been at work researchers have been concerned with quantifying the amount of load, or physical stress, placed on him. Advances in automation and technology however, have recently changed the nature of man's work from that of physical laborer to mental laborer, shifting the primary focus from the human's physical capabilities to the level of cognitive, or mental load with which the human can effectively cope. Estimation of a worker's ability to handle a mental task has revealed itself to be a more complex undertaking than the analogy originally suggested.

Many techniques have been used, some successfully and some not as successfully, in the effort to determine the nature and extent of the cost to the human operator for performing cognitive work. In general, methods can be classified into three broad categories, most of which will be addressed in this paper. The categories are: performance measures, subjective measures, and physiological measures. Performance measures assume that the operator's interactions with the system will result in different levels of performance depending on the difficulty of the task. Thus, such measures reflect whether or not the operator is able to meet the demands of the task. Increased task difficulty will manifest itself in the form of increased errors and slower reaction times. Unless secondary task methodology is used, however, these measures do not provide any indication of how much spare capacity the operator may have to perform additional tasks.

Subjective measures are based on the assumption that an operator is able to evaluate his own level of workload and thus these measures utilize a set of questionnaires on which the operator rates his degree of load. In addition to being convenient, subjective techniques are diagnostic, and often reveal sources of workload attributable to an operator's internal characteristics such as motivation, frustration, etc.

Physiological measures are based on the premise that mental tasks are performed at a certain physiological cost to the operator, with indications of load showing up in a number of observable physiological systems. The list of indicators is long, and includes measures of heart rate, heart rate variability, respiratory activity, blood pressure, body

temperature, galvanic skin response, direction of eye movements, urochemical analysis, pupil diameter, muscle tension, and event-related cortical activity (ERP's). The most obvious advantages of the physiological measures over the rest are their relative objectivity, their ability to be recorded continuously, and their unobtrusivity in operational settings. Since the greater portion of workload research being done today is directed at the operator at work (pilots, in particular), the unobtrusivity of these measures stands out as one of their most attractive features. Of popular interest are the measures of cardiac functioning, which will be the focus of this paper.

Mental workload has shown to be a multidimensional construct reflecting the interaction of many factors, including an operator's training and skill level, task demands, as well as the operator's physiological state, which itself is a function of manifold homeostatic systems. To prove reliable, an approach to mental workload estimation must be malleable to the dynamic nature of the concept of workload itself.

As an example, suppose I wished to evaluate the level of frustration of a subject performing a difficult versus an easy war-type video game. Further, suppose that I employed two different dependent variables - number of enemy "hits" and heart rate. When the results of the "experiment" are analyzed I find that the difficult game produces a much higher heart rate in the subject than does the easy game, but the number of hits is the same for the two. This point illustrates the fact that different measurement devices are sensitive to different components of workload - physiological measures tap operator strain or effort (not to mention physical load), and performance measures reflect on the difficulty of the task. It may very well be that the two games were both too easy or both too hard, revealed by the fact that performance was the same on both. Nonetheless, the performance measure has told me nothing of the subject's level of frustration during the two tasks.

In the search for measures useful both in the laboratory and in operational environments it is highly unlikely that one approach, or measuring stick, will provide all the answers, since what is being measured is a dynamic and multifaceted concept. Careful definitions of mental workload paired with careful selection and implementation of a number of metrics is currently the most promising of steps toward a solution. Since the rigors of defining mental workload have been covered elsewhere in this volume, this effort will focus on a review of several approaches to the study of mental load using cardiac measures, and on the combination and interpretation of several metrics from different classes in a divided attention task performed in the laboratory.

Relationship of physiological systems to cognitive systems

According to Hancock (ref. 1), "If ERP's represent the highest scoring physiological measure on the scale of spatial and

systemic congruence with respect to CNS activity, then measures pertaining to heart rate and its derivatives are currently the most practical method of assessing imposed mental workload".

Before beginning a review of studies employing cardiac measures of load there are several important issues that need be addressed. The first of these major questions facing the scientist using physiological measures of cognitive processing concerns the exact relationship between the physiological systems and the cognitive systems. The term system is used here to represent a highly complex inter-connected network of processes that are constantly changing and approaching a goal that is oftentimes unknown. How do the physiological systems respond to different levels of cognitive processing? Is there really a physiological cost to thinking? Although perhaps more obvious to those using physiological measures, the relationship problem is nonetheless present in every approach to quantifying workload.

A widely-held biological conception is that the physiological processes are in constant oscillation seeking a homeostatic state that will balance input from environmental factors, self-generated information, task-specific information, and biological functioning (refs. 2 and 3). The forecast for someone trying to measure the physiological cost associated with varying levels of cognitive load is grim from this perspective, since the physiological systems are "programmed" towards homeostasis and will adjust what parameters are necessary to keep things in even keel. It is possible that overall system output could remain the same due to the operator not performing a required task, or by the adoption of strategies altering the level of performance of several tasks. The physiological system keeps itself in a state of preparedness for emergencies by storing a certain level of "reserve capacity" to be used only in extreme cases (ref. 3). Situations most likely to allow use of the reserve capacity include extremely fearful or stressful situations, extreme physical loads, extremes of temperature, etc. These are not the situations normally encountered in a laboratory experiment; thus, ~~if Kalsbeek's notion is correct~~, few studies should show physiological correlates of mental load. A quick glance through the literature will show that this is not the case. Many studies report changes in physiological processes associated with manipulated changes in mental load. Unfortunately, the problem is quite the opposite - the influence of too many variables is evident in cardiac records. One technique, however, the spectral decomposition of the heart inter-beat interval into its constituent frequency components, shows the most promise for looking at, if not unconfounding, the variances associated with a number of different physiological systems. This promising avenue will be explored later in this report.

Factors associated with cardiac output

Once one is willing to accept the idea that physiological

processes are an accurate reflection of implicit mental processing, one must also realize that cardiac functions are also affected by a number of factors not thusfar known to be related to cognition. Documented correlates include age, temperature, emotions, physical load, level of responsibility, level of task-related risk, respiration, and noise (refs. 4 and 5). Even in the most carefully conducted laboratory experiment many of these factors are difficult, if not impossible to control. The state of affairs worsens as one considers the current interest in applying measures of workload in operational environments where even less control is possible.

Grain of analysis

As with other measures of workload, an issue of debate is the unit of measurement, or grain of analysis used in recording and summarizing data. Research has shown that different results may be found depending on whether data (reaction time, d') are averaged over all of the trials within a block or conditional upon the types of trials comprising a block (only one response required, two responses required) (refs. 6 and 7). The three measures to be discussed in this paper differ in the amount of data that is collapsed over, with mean heart rate spanning the most, followed by overall heart rate variability, followed lastly by spectral analysis. A number of researchers have expressed concern over studies reporting data based on summary statistics for heart rate data inherently based on a non-random time series (refs. 8 and 9).

Related to the grain of analysis problem is the issue of whether cardiac responses to levels of tasks or to components of tasks should be observed (ref. 4). Should data be averaged over a block of trials of the same task (e.g. difficult mental arithmetic vs easy mental arithmetic) or over similar parts of a task occurring across trials (e.g. stimulus perception, mental rotation, etc.)? Clearly, those interested in operator responses to overall levels of mental load (that is, ergonomists) are interested in the first question. Any indicator sensitive to varying levels of task load is useful to someone with that purpose in mind. But to the cognitive psychologist, who is interested in discovering the architecture of the processing system, the second alternative appears more attractive. Ultimately, all researchers, basic and applied, are interested in a priori prediction of workload levels given certain task combinations. Thus, the major problem has two parts. A detailed analysis of laboratory tasks used in workload studies must be first undertaken, so that the components comprising a given task may be clearly specified. This would be followed by examination of cardiac responses associated with each component (e.g. perceptual input, central processing, and response processing) of the task. Only then can predictions be made concerning workload levels inherent in untested combinations of the examined task components.

The next sections will present a critical review of several

studies using each of the cardiac measures of workload - mean heart rate, overall heart rate variability, and spectral analysis of heart rate.

Mean heart rate

Unless stated otherwise, it is assumed that HR is measured offline. Although there are some recent developments in online measurement techniques, (ref. 10) most research reports data that were collected as interbeat interval scores and subsequently analyzed offline, although ECG's provide a visual report of the data during the experiment (ref. 11).

As mentioned previously, mean HR makes the least parsimonious use of the available heart inter-beat interval data of the three measures. The overall statistic of HR is computed as $1/IBI$ (in seconds). Most studies using mean HR as a dependent variable take an average of the HR over each task period or experimental condition. Some studies, however, report second-by-second levels of mean HR (collapsed across trials and subjects) so that an approximation of the complete waveform may be seen. Such an approach is to be preferred to condition means since it is known that HR is extremely variable during the first few seconds of a task and may contaminate the data from the rest of the recording interval. Plots of the overall trend can be observed and outlying data removed from subsequent analysis.

Lacey's intake-rejection hypothesis

The majority of experiments reviewed were directed at supporting or providing evidence against Lacey's intake-rejection hypothesis (ref. 12). Specifically, Lacey proposes that an acceleration in HR accompanies tasks requiring complex "internal" processing such as mental arithmetic or memory scanning. Accordingly, HR deceleration accompanies tasks requiring attention or responses to external stimuli. The cardiovascular system is presumed to exert an influence on the bulbar-inhibitory area of the brain, which serves to enhance or inhibit detection of sensory inputs. Such responses are said to be biologically adaptive in that a faster HR is effective in shutting out potentially distracting noise so that the internal processing may proceed unhindered. HR deceleration supposedly reduces internal noise, enhancing signal detection sensitivity. Such a process would result in faster reaction times and increased accuracy to stimuli.

In the earliest of the reviewed studies addressing the intake-rejection hypothesis, Kahneman, Turskey, Shapiro, & Crider (ref. 13) observed mean HR, pupil diameter, and skin resistance to phases of a task in which subjects added 0, 1, or 3 to each of 4 serially-presented digits, and reported the transformed series. Although task difficulty effects were seen only in the skin resistance and pupillary measures, all measures reflected an increase in the phase of the task where the digits were mentally

manipulated, followed by a peak and sharp decline in the response phase, supporting Lacey's hypothesis. Problematic for the experiment is a trend towards differences in the dependent variables among the three levels of difficulty conditions prior to any procedural differences in the tasks (i.e. prior to digit presentation).

In a more common manipulation of attentional direction, Coles (ref. 14) instructed subjects to search a 40 x 60 letter array for targets either highly discriminable or not easily discriminable from the background letters. The targets were the letter "e" or the letter "b", distributed with varying density among the letter "a" distractors. Detected targets were either counted (internally-directed attention) or denoted by a check mark (externally-directed attention). Support for Lacey's hypothesis was found, since decreased target letter discriminability resulted in decreased HR (and increased HR deceleration), and counting targets caused HR to decelerate while checking targets caused HR to accelerate. As with the Kahneman et al. (ref. 13) experiment, pre-search task differences in mean HR for the two search conditions overshadowed the findings, not to mention the fact that physical workload was also greater in the externally-directed attention condition where the subjects checked each target detected. Also, complete testing of Lacey's hypothesis was not possible due to the unavailability of reaction time data (except in the form of # of lines searched) in the task. As mentioned previously, decreased HR producing enhanced sensitivity for externally-presented stimuli should be reflected in reaction time and accuracy in the task. No error data was reported in the study.

The major argument for an alternative explanation of cardiac acceleratory and deceleratory changes involves the level of verbalization involved in the tasks (ref. 15). Presumably, "intake" tasks are associated with a higher level of internal verbalization than are "rejection" tasks. Klinger, Gregoire, & Barta (ref. 16) measured mean HR, rapid eye movements (REM's), and electroencephalogram alpha levels (EEG) in tasks where subjects performed mental arithmetic, counted aloud by two's, indicated preferences between two activities, mentally searched among alternatives, imagined a liked person, or suppressed thoughts of a liked person. The levels of HR found in the study were, from highest to lowest, in the order of the tasks just given. Tasks associated with the three highest levels of HR involved both concentration (internal processing, or rejection tasks, according to Lacey) and verbalization. Thus there appears to be a plausible (and more parsimonious, according to some) explanation for the observed set of data.

15. → (ref. 15)
Elliott has criticized Lacey's intake-rejection hypothesis and studies supporting it. Besides claiming that there is a general lack of empirical support for the hypothesis, (a disputable claim, upon surveying the literature) he further argues that the hypothesis is untestable due to the lack of sufficient operational definitions. A more parsimonious account,

he suggests, is Obrist's conception of a cardiac-somatic relationship (ref. 17), where HR changes are attributed to motor activity. In this sense, HR is used as a response, and not as a cause of changes in processing efficiency. This leads the discussion to the arousal model, to be reviewed next.

Arousal models versus mental load models

The Yerkes-Dodson Law predicts an inverted U-shaped function relating performance on a mental task to the level of arousal, or stress befalling the performer. Zwaga (ref. 18) argues that the concept of arousal is a better account of observed HR changes during an experiment. Zwaga gave his subjects a paced mental arithmetic task consisting of five minutes of rest, six minutes of the arithmetic task, and five more minutes of rest. Heart rate during the first minute of the task was the highest, and thus was discarded. He further found that HR during the task was higher than that during the rest periods, but that HR decreased with the duration of the task period. HR also declined with each session of the experiment, even when the sessions were separated by a 24 hour period. Although a mental load model would predict higher HR during the task period than in rest, such a model has no explanation for why HR continued to decrease throughout the task period and with further sessions. Such findings are easily accommodated by an arousal model that predicts eventual habituation to repeated presentations of stimuli.

Cacioppo & Sandman (ref. 19) maintain that the level of cognitive demands of a task, and not a general level of sympathetic arousal, are the reason underlying observed HR effects. In their experiment, subjects were given either problems to solve (anagrams, arithmetic, or digit-string memorization), or slides of autopsies to look at. The autopsy slides were associated with two levels of stressfulness, with low stress slides being pictures taken from a distance of an accident victim, and high stress slides being close-ups of badly-mutilated accident victims. The assumption was made that stressfulness was equivalent to unpleasantness, with difficult cognitive tasks being rated as more unpleasant or stressful than easy cognitive tasks. Measuring only the first five heart beats in each task condition, difficult (stressful) cognitive tasks were associated with higher HR than easy cognitive tasks, while the stressfulness of the autopsy slides did not affect HR. Averaging over difficulty, cognitive tasks produced an increase in HR, while autopsy slide viewing produced a decrease. An arousal hypothesis would have predicted increased generalized sympathetic responses to the stressful autopsy slides relative to the low stress slides, and increased overall HR to the autopsy slides relative to the cognitive tasks. Since this was not found the authors concluded that mental processing demands associated with cognitive tasks are responsible for observed HR changes. The conflict between the two competing hypotheses could possibly be resolved by equating the measurement procedures (discarding obviously outlying HR scores obtained in the first few minutes of

a session).

Laboratory versus field findings

Two of the reviewed experiments observed HR in operational environments, and found virtually no changes associated with mental load. This finding is surprising compared with the wealth of evidence supporting the use of HR to measure mental load in the laboratory. Melton, Smith, McKenzie, Wicks, & Saldivar (ref. 20) studied mean HR, urine steroid, epinephrine, and norepinephrine levels, and level of anxiety in air traffic control (ATC) workers employed at low traffic control centers. In contrast to findings of studies at high-density traffic centers, no HR increases from off duty to on duty were observed in the ATC workers.

A comprehensive study evaluating 20 different workload measures, including HR and heart rate variability (HRV), was conducted by Wierwille & Connor (ref. 21) using a simulator in three levels of flight difficulty. Of the physiological measures studied, only mean pulse rate was observed to increase monotonically with imposed flight difficulty. No effects on HRV (scored by the standard deviation) were observed. Subjective measures, followed by performance measures, were the most sensitive to imposed load.

Hart & Hauser (ref. 22) found that the level of pilot responsibility (left seat versus right seat) and the segment of flight were able to produce changes in mean HR. HR was higher for the pilot in control of the plane than for the co-pilot, and was higher during take-off and landing phases segments compared to segments of level flight. A major problem with field studies, even if observed changes in HR are observed, is the lack of environmental control. A useful distinction among types of stress has been suggested, and that is the consideration of informational versus emotional stress. Presumably an operational environment, especially in flight, would contain more levels of emotional stress than that encountered in a laboratory, while informational stress could potentially be the same in the two environments. An experiment by Sekiguchi, Handa, Gotoh, Kurihara, Nagasawa, & Kuroda (ref. 23) in which six tasks were used ranging from tracking in the laboratory to an actual flight task supported such a notion. Perhaps the arousal hypotheses, although not useful in the laboratory environment, holds potential for testing in operational environments.

Heart rate variability

The major problems facing researchers using heart rate variability, or sinus arrhythmia, as a dependent measure are associated with 1) the choice of a valid and sensitive scoring method, and 2) how to remove (or prevent) contamination of observed results by influences unrelated to cognitive processing, e.g. physical load, respiration, etc.

Data scoring

Statistics used to estimate the degree of variability among a collection of IBI scores include the typical standard deviation, the number of reversals (points of inflection) in the HR signal (ref. 24), the frequency that the HR signal crosses the mean or 3, 6, or 9 beats per minute on either side of the mean (ref. 25), and the mean square of successive positive or negative (or both) differences (MSSD) between the heart rate signal. Essentially, the various scoring methods differ as to how much data is collapsed over, and whether amplitude or frequency information is included in the calculation. A comprehensive review of factor and spectral analytic techniques is provided by Opmeer (ref. 26).

Since so many empirical factors are allowed to vary, even when the selection of a scoring method is held constant, no particular statistic emerges as best in any given situation. There is some indication, as will be discussed in the section on spectral analysis, that those methods accounting for the direction and amplitude of change in the IBI are the most sensitive.

Physical versus mental load

It has been typically observed that increases in imposed physical load elevate mean HR while increases in imposed mental load decrease HRV. Such effects have often been obscured, however, due to the employment of a binary choice task at differing rates of stimulus presentation as a manipulation of task difficulty. Such a treatment confounds levels of mental load with levels of physical load. Unfortunately in some cases this confound can "cancel out" HRV effects actually due to increased mental load. Kalsbeek & Sykes (ref. 27) used such a procedure and failed to find HRV differences between levels of task difficulty.

In a classic study, Boyce (ref. 28) factorially manipulated levels of physical and mental load in an attempt to separate effects on HRV (measured by the standard deviation) associated with the two factors. Subjects were given a one- versus two-digit mental arithmetic task in which they had to move a pointer (attached via a cable to a weight) to the correct answer. Physical load was varied by changing the heaviness of the weight attached to the end of the cable. Results indicated an increase in mean HR due to both physical and mental load, while HRV decreased with increases in mental load and increased with increases in physical load.

Inomata (ref. 29) found no HR or HRV differences among rest periods and periods of a visual search task characterized by four levels of memory load, and no differences between those measures among the four load conditions. HRV was scored using the standard deviation and the sum of the frequencies per minute

crossing the mean or 3, 6, or 9 beats per minute away from the mean. When the data were re-analyzed after removing data associated with overt body movement (subject's moving in their chairs, etc.), only the second deviation score decreased with increasing memory load.

Using a more complex statistic, Luczak (ref. 30) gave subjects a binary choice reaction time task with and without physical load. HRV was scored as the ARQ, ~~which was equivalent to dividing all of the positive differences (in rate) between successive heart beats by the frequency of relative maxima and minima in the time series.~~ Physical load was achieved by having subjects move various parts of their body at the same time as they performed the binary choice task. They found that HR was correlated highly with motor load, while HRV was correlated with mental load. HRV decreased with increasing task difficulty.

Despite a confound with physical load, Ettema & Zielhuis (ref. 25) found increased HR, blood pressure, and respiration and decreased HRV with increasing levels of mental load achieved using a paced binary choice task at 20, 30, 40, and 50 signals per minute. The heart rate, blood pressure, and respiration measures were all positively correlated with each other, and negatively correlated with both measures of HRV. HRV was scored as either the frequency of HR above or below 3, 6, or 9 beats away from the mean, or as the sum of the absolute differences between successive levels of HR.

Spectral analysis of heart rate variability

Unlike the two methods just discussed, which focus on the overall variability of the cardiac signal, the spectral analysis technique treats the IBI data as a time series upon which analysis methods in the frequency domain or the time domain may be applied. Debate has arisen concerning the appropriateness of using the typical analysis of variance statistics, which assume random samples, on non-random data. Specifically, Luczak & Laurig (ref. 8) have pointed out that when such statistics are used on time series data of IBI's the degrees of freedom associated with the experimental conditions are overestimated. This is because the samples are not random and reflect the interaction of many rhythmically occurring functions in the autonomic nervous system. It is obvious to most that the overall mean or variance of such a series does not reflect the rhythmicity of the underlying processes. Two alternative procedures remain: analysis methods from the time domain, and analysis methods from the frequency domain.

Time domain methods

Methods in this class involve the shifting of a time series in time by a specified amount of lag, and then either correlating the signal with itself (autocorrelation) or with another series

(cross-correlation), in order to see power trends in the data. Since there is a great deal of noise present in the series, noise that is usually not of empirical interest, it must be removed before the factors of interest can be examined. Noise removal techniques are complex and are discussed in further detail in Coles et al. (ref. 31). In general, time domain methods have been left to scientists in electrical engineering, with psychologists choosing to employ more traditional analysis techniques.

Frequency domain methods

Analysis of heart rate variability in the frequency domain shows the greatest promise among all the cardiac measures as a reliable indicator of operator workload. Despite its methodological and theoretical promise, fewer papers have been published using this method than the two previously discussed, no doubt due to its greater complexity. These techniques, known as spectral analysis, or harmonic analysis, break the cardiac signal down into its constituent frequency components. Conceptually this is similar to the way total variance is partitioned into that accounted for by main effects and interactions in an analysis of variance (ref. 9). First, the series is transformed into one sampled at equal intervals (since most data is a measure of the R-R interval, which varies), and then a Fourier analysis is performed which reveals the amplitude of the variance at each frequency of the signal. The sum of the energies in each interval is equal to the overall variance of the IBI. Partitioning the variance, or energy, in this way allows the researcher to see the effects of a manipulation on the individual components of the cardiac signal, even if those effects can't be controlled for in the first place. Although it is considered a more elegant technique than the others, use of the technique alone is no substitute for careful experimental design to minimize influences from sources other than those of interest. Experiments should be designed to minimize potential confounds from rhythmically-occurring biological processes that are not specifically related to cognitive processing per se., such as the time of day, ambient temperature, etc.

Different biological functions contribute power to different frequencies of the total cardiac output. The results from experiments using spectral analysis of IBI data usually reflect a body temperature component at about 0.05 Hz, a blood pressure component around 0.1 Hz, and a respiratory component in the area between 0.25 and 0.40 Hz, the normal adult breathing rate of 15 - 24 breaths per minute (ref. 32). In addition, a component may appear around the same frequency as the task presentation rate. If the task were a binary choice task with stimuli presented once every 2 seconds, a task-related component may occur at 0.5 Hz. Such a phenomenon has been called "entrainment", and refers to the synchronization of certain internal rhythms with external ones. The effect arises due to HR deceleration just prior to an expected stimulus, and acceleration just after stimulus

presentation. There is also evidence that blood pressure can be entrained by respiration if the respiration rate is high and deep (ref. 33).

Not all researchers have shown the same degree of concern for the influences of respiration on the distribution of power in the cardiac spectrum. Mulder & Mulder (ref. 32) intentionally manipulated subjects' frequency and depth of respiration alone and while engaged in cognitive tasks. Results indicated that frequency bands toward the low end of the spectrum (e.g. 0.06-0.14 Hz) were not at all affected by respiration, while moving up the spectrum found effects of both frequency and depth. Increasing the difficulty of cognitive tasks was found to decrease the power inherent in a frequency band around 0.1 Hz relative to other frequency bands. Mulder & Mulder described the power at 0.1 Hz as an indicator of the amount of time spent in "controlled processing".

Spectral techniques have also been used in environments other than the laboratory. One study used tasks ranging from bedrest to treadmill exercise to tracking and actual flight that showed the power in the 0.1 Hz range to increase with moderate mental load, and decrease with increases in mental load (ref. 34). In the flight task, power in the .1 Hz range increased in the preflight check and decreased during takeoff and landing, a result complimented by HR studies (ref. 22).

One operational environment in particular, however, has turned up results contrary to those found in flight environments. Egelund (ref. 35) reports that most studies of driving find that HR decreases with the number of hours driven, while HRV tends to increase, presumably due to fatigue. The physical work associated with maneuvering a vehicle in traffic contributes to increases in HR. Nygaard and Schiotz (ref. 36) had subjects drive a 340 kilometer course on either straight flat highways or ones with many hills and turns. They found no difference in HRV (as measured by single deviant heartbeats) between the two types of roads. Suspecting insensitivity of their measure, among other factors, Egelund (ref. 35) re-analyzed Nygaard and Schiotz's data using spectral analysis of the interbeat interval data, HRV (the standard deviation), and mean HR. Egelund predicted that the 0.1 Hz region of the spectrum would reflect an increase over the amount of time driven, while HR would decrease over time. No changes in HR or HRV were found as a function of distance driven, however, a slightly significant increase in the variability in the 0.1 Hz region was found for 2 of the last 5 segments of the journey. Although the results supported those from an earlier study, their statistical weakness was blamed on a number of factors, namely, the shortness of the test drive, and driver experience. It is worthy to note that 4 of the 8 subjects had had their licenses for two and one-half years or less (one had even had hers for only 2 weeks).

Earlier in this paper some of the problems associated with using the usual summary statistics on time series data were

mentioned. A possible solution to this problem has materialized in the form of a summary statistic appropriate for spectral analytic techniques, called the weighted coherence (ref. 9). The statistic is useful for correlating the power variations at one frequency with those at another. This would allow the power variability at the respiratory frequency to be correlated to the variability at the 0.1 Hz frequency, for example. Currently it is possible to do a cross-spectral analysis, where the coherence (similar to r^2) of one rhythm with another at one specific frequency can be determined. However, without prior knowledge of which exact frequencies are of interest it was not possible to get this statistic to apply to a range of frequencies. The proposed measure, the weighted coherence, is an indication of the total variance shared by two rhythms within a limited frequency band. Finally, a means of summarizing across frequencies is available, although Porges and his colleagues did not report data validating the statistic.

The divided attention experiment

Next we will report on an experiment carried out in our laboratory combining performance and physiological measures of workload. Since the data were only recently collected, the findings reported are preliminary and much work remains to be done.

The task employed was a bimodal divided attention task in which subjects simultaneously attended to two streams of discrete stimuli, and responded manually to changes in one modality and vocally to changes in the other modality. The events in the auditory modality were high or low-frequency tones lasting 100 msec, with 1100 msec allowed for response after tone presentation. The visual events were 100 msec flashes of a red or green light, with the same response interval as for the auditory task. A sequence of events lasted for 160 trials, or about 3.2 minutes. Subjects were instructed to respond as quickly as possible via either a keypress or by saying the word "diff" into a microphone, each time they observed a signal in a modality that was different from the previous signal in that modality. Half of the subjects used a vocal response to the auditory channel and a manual response to the visual channel, while for the other half of the subjects the response requirements were reversed. It should be noted that the response mappings for the former group should lead to better performance, since input and output modalities are more compatible for the auditory task than those used by the latter group (ref. 37). Tasks employing multiple modalities are useful in that they parallel tasks in operational environments more than the more traditional laboratory tasks, both in their difficulty and in their multimodal nature.

Task difficulty was manipulated by varying the number of tasks simultaneously performed (one = single stimulation, two =

double stimulation), and the degree of synchrony between two tasks. In the synchronous case, the auditory and visual stimuli occurred simultaneously, with a total of 1100 msec allowed for the subject to respond to both of the tasks. In the asynchronous case, presentation of the auditory or the visual sequence was delayed by 300 msec after that in the other modality. Presumably, tasks that occur asynchronously in each modality are easier to perform since attention may be switched between the two and responses need not necessarily be executed simultaneously.

Dependent variables were reaction time (RT), d' and beta (response criterion), and heart rate. For the first three measures, the data were examined both on an overall basis, and conditional upon the type of trial in the other modality: no response, response. Several cardiac measures were calculated, including mean HR, HR variance, mean successive differences in HR, variance of successive differences in HR, and the variability in the .1 Hz region of the power spectrum.

Performance measures

Not surprisingly, RT reliably distinguished between the easy and difficult levels of the task, with scores being fastest during single stimulation, and slowest during double stimulation. There is no a priori reason to suspect a difference in RT's between the auditory lagged and the visual lagged conditions, and there was none found. In general, as has been previously found, RT's to the visual channel were faster than those to the auditory channel. The visual RT advantage was most evident during the easier (one task lagged) versions of the task than during the more difficult task where auditory and visual stimuli were presented simultaneously. Subjects responded more quickly with practice, and were faster when the response modalities were compatibly arranged than when incompatibly arranged.

D' scores were not significantly different in the easy and difficult versions of the task, although the trend was in the right direction, with d' slightly higher in the easy condition. Contrary to the RT results, d' was higher for the auditory than for the visual channel, however the pattern was the same as the RT results with the auditory d' advantage being greater during the asynchronous tasks than during the synchronous task. A compatible response modality for the auditory channel also produced higher d' scores than the incompatible arrangement. Given conflicting RT and d' results we intend to examine the reaction time density functions to see if the response for one modality was always executed before that to another modality, or if sometimes the response order traded off between the two modalities. Such data should reveal whether capacity was shared between the two (dependent processes) or re-allocated to the other task once a task was completed (independent processes).

Values of beta were lowest in the synchronous condition, and more comparable between the two asynchronous conditions. Beta was also highest for whichever modality used a vocal response.

This measure is useful in distinguishing increased performance from merely a lowered subjective criterion to respond, as opposed to a true increased sensitivity to the signal events. As was expected, the most difficult condition, the synchronous condition, resulted in the lowest values of d' (although not significant), paired with the lowest values of β , indicating that even though the criterion to respond was lowered the subjects could still not effectively distinguish the signals from the noise.

Previous experiments in this series have shown there to be an asymmetric trade-off of performance between the auditory and the visual channels dependent on whether or not 1) a response is made in the other channel, and 2) whether or not that response is overt (hit) or implicit (correct rejection) (ref. 7). Performance in the auditory channel is best when there is no overt response made to the visual channel, and worst when there is an overt response to the visual channel. Performance in the visual channel has not been shown to be affected by events in the auditory channel, for reasons beyond the scope of this paper. Further breakdowns of the data show that the visual response events causing the auditory performance decrement are both hits and false alarms, implicating interference between the channels at the response stages of processing.

At the present time we are able to report data for RT conditioned on whether or not there was a response in the opposite channel. RT was significantly faster when no response (either a hit or a false alarm) was executed in the opposite channel. The interaction of trial type with modality revealed that the RT advantage on no response trials was shown only for the visual channel. The frequency differences between the high and low tones are suspect for causing this apparent departure from earlier findings.

Cardiac measures

At the time of this report, HR data was available for 6 of the 24 subjects run in the experiment. Mean HR scores showed a decrease in HR throughout the experiment. Of HR, HR variance, mean successive difference in IBI's (MSD), and variance of successive difference in IBI's, only mean HR reflected differences between the pre-task baseline period (82 BPM) and the task period (76 BPM). HR did not distinguish, however, between the single and double stimulation versions of the task.

HR variance was significantly greater during the last half of the experiment than in the first half, but decreased within a half, perhaps reflecting the fact that subjects were growing increasingly fatigued and exerting greater effort during the portions of the experiment between rest periods.

Although not significant, the MSD measure was positive (reflecting decelerating HR) during the baseline period and

negative (reflecting accelerating HR) during the task period. MSD variance did not show any effects of any of the experimental manipulations.

The IBI data were subject to interpolation to create a regularly-sampled sequence, and were input to a spectral analysis program revealing the density at each frequency in the spectrum. The power in four different frequency bands was examined: 0.06-0.14 Hz, 0.16-0.24 Hz, 0.26-0.32 Hz, and 0.34-0.42 Hz (ref. 32). Analysis of variance did not reveal differential sensitivity of the four frequency bands to manipulations of task difficulty. Several factors may account for the null findings. Although it seems plausible that our divided attention task should be at least as difficult as those reported previously using HRV as a measure, it is possible that it was not so difficult as to cause differing degrees of effort in the subjects. No performance criteria were imposed on the subjects, resulting in a higher than average number of missed responses and false alarms. The signal detection measures rely on the assumption that humans are less-than perfect observers, so performance errors were not discouraged. Another possibility relates to the way the analyses were performed. Power within a band was averaged over several frequencies, possibly canceling out any effects. Mulder (ref. 38) reported data separated into discrete frequencies that showed that the 0.06 and 0.08 frequencies in particular were the most sensitive to task difficulty. Further breakdowns of the data should either support or rule out such an interpretation, which will have to be regarded as speculation until then. Not to be excluded from consideration is the fact that 3/4 of the heart rate data has not yet been analyzed, implicating insufficient power in the present null results.

Future experiments will also examine phasic HR, in a manner similar to the experiments reported earlier by Kahneman et al. (ref. 13) and Coles (ref. 14). The divided attention task has potential as a task using longer trials such that cardiac responses during different segments of a trial may be observed.

General conclusions

The importance of addressing of mental workload as a multidimensional construct is cannot be overemphasized. The potential for interactions among metrics used to assess load and the degree of imposed load is great and oftentimes unpredictable. The importance of two factors is evident: careful experimental design, and a grain of data analysis appropriate to the characteristics of the monitored signal.

Separating overall variability into smaller parcels allows us to observe the interrelationships among the different biological systems as they are related to mental processing. For physiological systems at least, the closer the data resembles continuous data, the better. At this point it seems clear that even though apparently extraneous influences can be observed and

documented, they cannot be removed. Since a human is a complex system, complex responses to external and internal demands will be reflected in empirical data. Spectral analytic techniques are extremely powerful and useful tools for assessing external attentional demands placed on operators, but use of them will not guarantee solution of the workload evaluation problem. No matter what degree of experimental control is exercised over an experiment, the operator at work is going to be under a number of uncontrolled, and perhaps even unknown influences, all of which interact dynamically to result in a given level of operator strain. Nonetheless, fractionization of the task components, as well as the associated measures of workload and performance, appears to be the surest path to the study of understanding the nature of the interaction.

ACKNOWLEDGMENTS

This research was supported by Cooperative Agreement NCC 2-228 from the National Aeronautics and Space Administration, Ames Research Center; S. G. Hart was the Technical Monitor.

REFERENCES

1. Hancock, P. A. (1985) Physiological reflections of mental workload. *Aviation, Space, and Environmental Medicine*, 56, 1110-1114.
2. Cannon, W. B. (1932) *The wisdom of the body*. New York : W. W. Norton.
3. Kalsbeek, J. W. H. (1973) Do you believe in sinus arrhythmia? *Ergonomics*, 16(1), 99-104.
4. Firth, P. A. (1973) Psychological factors influencing the relationship between cardiac arrhythmia and mental load. *Ergonomics*, 16(1), 5-16.
5. Hart, S. G. (1986, ~~in press~~) Theory and measurement of human workload. In J. Zeidner (Ed.), *Human Productivity Enhancement*, New York : Praeger.
: *Training and Human Factors in System Design*, Vol. II (pp. 392-432).
6. Casper, P.A. & Kantowitz, B. H. (1985) Seeing tones and hearing rectangles: Attending to simultaneous auditory and visual events. *Trends in Ergonomics/Human Factors II*, in R. E. Eberts & C. G. Eberts (Eds.), Elsevier Science Publishers.
7. Casper, P.A. (1986) A signal detection analysis of bimodal attention: Support for response interference. Unpublished Master's Thesis. Purdue University.
8. Luczak, H., & Laurig, W. (1973) An analysis of heart rate variability. *Ergonomics*, 16(1), 85-97.

9. Porges, S. W., Bohrer, R. E., Cheung, M. N., Drasgow, F., McCabe, P. M., & Keren, G. (1980) New time series statistic for detecting rhythmic co-occurrence in the frequency domain : The weighted coherence and its application to psychophysiological research. *Psychological Bulletin*, 88(3), 580-587.
10. Adie, P., & Drasic, C. (1986) Validation of a mental workload measurement device. Unpublished master's thesis. Department of Industrial Engineering, University of Toronto.
11. deBoer, R.W., Karemaker, J.M., & Strackee, J. (1985) Description of heart-rate variability data in accordance with a physiological model for the genesis of heartbeats. *Psychophysiology*, 22(2), 147-155.
12. Lacey, J. I. (1967) Somatic response patterning and stress: Some revisions of activation theory. In M. A. Appleby & R. Trumbull (Eds.), *Psychological stress: Issues in research*, New York: Appleton-Century-Crofts.
13. Kahneman, D., Turskey, B., Shapiro, D., & Crider, A. (1969) Pupillary, heart rate, and skin resistance changes during a mental task. *Journal of Experimental Psychology*, 79(1), 164-167.
14. Coles, M. G. H. (1972) Cardiac and respiratory activity during visual search. *Journal of Experimental Psychology*, 96(2), 371-379.
15. Elliott, R. (1972) The significance of heart rate for behavior : A critique of Lacey's hypothesis. *Journal of Personality and Social Psychology*, 22(3), 398-409.
16. Klinger, E., Gregoire, C., & Barta, S. (1973) Physiological correlates of mental activity : Eye movements, alpha, and heart rate during imagining, suppression, concentration, search and choice. *Psychophysiology*, 10(5), 471-477.
17. Obrist, P. A., Webb, R. A., Sutterer, J. R., & Howard, J. L. The cardiac-somatic relationship : Some reformulations. *Psychophysiology*, 6, 569-587.
18. Zwaga, H. H. G. (1973) Psychophysiological reactions to mental tasks : Effort or stress? *Ergonomics*, 16(1), 61-67.
19. Cacioppo, J. T., & Sandman, C. A. (1978) Physiological differentiation of sensory and cognitive tasks as a function of warning, processing demands, and reported unpleasantness. *Biological Psychology*, 6(3), 181-192.

20. Melton, C. E., Smith, R. C., McKenzie, J. M., Wicks, S. M., & Saldivar, J. T. (1978) Stress in air traffic personnel : Low-density towers and flight service stations. Aviation, Space, and Environmental Medicine, 49(5), 724-728.
21. Wierwille, W. W., & Connor, S. A. (1983) Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator. Human Factors, 25(1), 1-16.
22. Hart, S. G., & Hauser, J. R. (198⁷~~8~~, in press) Inflight application of three pilot workload measurement techniques. ~~Manuscript accepted for publication in~~ Aviation, Space, and Environmental Medicine. 58 402-411.
23. Sekiguchi, C., Handa, Y., Gotoh, M., Kurihara, Y., Nagasawa, A., & Kuroda, I. (1978) Evaluation method of mental workload under flight conditions. Aviation, Space, and Environmental Medicine, 49(7), 920-925.
24. Mulder, G., & Mulder-Hajonides van der Meulen, W. R. E. H. (1973) Mental load and the measurement of heart rate variability. Ergonomics, 16(1), 69-83.
25. Ettema, J. H., & Zielhuis, R. L. (1971) Physiological parameters of mental load. Ergonomics, 14(1), 136-144.
26. Opmeer, C. H. J. M. (1973) The information content of successive RR-Interval times in the ECG. Preliminary results using factor analysis and frequency analysis. Ergonomics, 16(1), 105-112.
27. Kalsbeek, J. W. H., & Sykes, R. N. (1967) Objective measurement of mental load. Acta Psychologica, 27, 253-261.
28. Boyce, P. R. (1974) Sinus arrhythmia as a measure of mental load. Ergonomics, 17(2), 177-183.
29. Inomata, O. (1977) An evaluation of heart rate variability in different levels of mental loading. Journal of Human Ergology, 6(2), 208-210.
30. Luczak, H. (1979) Fractioned heart rate variability. Part II : Experiments on superimposition of components of stress. Ergonomics, 22(12), 1315-1323.
31. Coles, M. G. H., Gratton, G., Kramer, A. F., Miller, G. A., and Porges, S. W. (1986) Principles of signal acquisition and analysis. In Psychophysiology : Systems, Processes, and Applications. Vol I : Systems. M. G. H. Coles, E. Donchin, and S. W. Porges (Eds.), New York : Guilford Press.
32. Mulder, G. & Mulder, L. J. M. (1981) Information processing

and cardiovascular control. Psychophysiology, 18(4), 392-402.

33. Sayers, B. McA. (1973) Analysis of heart rate variability. Ergonomics, 16(1), 17-32.
34. Sekiguchi, C., Handa, Y., Gotoh, M., Kurihara, Y., Nagasawa, A., & Kuroda, I. (1979) Frequency analysis of heart rate variability during flight conditions. Aviation, Space, and Environmental Medicine, 50(6), 625-634.
35. Egelund, N. (1982) Spectral analysis of heart rate variability as an indicator of driver fatigue. Ergonomics, 25(7), 663-672.
36. Nygaard, B., & Schiotz, I. (1975) Monotoni pa motorveje. Radet for Trafiksikkerhedsforskning, Notat 135, Kobenhaven.
37. Wickens, C. D. (1980) The structure of attentional resources. In R. S. Nickerson (Ed.), Attention and Performance VIII, Hillsdale, N.J.: Erlbaum.
38. Mulder, G. (1979) Sinusarrhythmia and mental work load. In N. Moray (Ed.), Mental Workload: Its Theory and Measurement. New York: Plenum Press.

ORIGINAL PAGE IS
OF POOR QUALITY

APPENDIX D:

Casper, P.A., Shively, R.J., & Hart, S.G. (1987) Decision
support for workload assessment: Introducing WC FIELDE.
Proceedings of the Human Factors Society 31st Annual Meeting,
72-76.